

GC10432d

Title: CLIMATE PREDICTABILITY OF EXTREME FLOODS IN THE UNITED STATES,
PI/coPI: Upmanu Lall, Yochanan Kushnir, Andrew Robertson, Jennie Nakamura
Report Year (Final Report), 2014
Grant # NA10OAR4310159

1. Results and Accomplishments

As indicated in our previous reports, we conducted research on an understanding of large scale climate drivers of extreme floods. In this context, our research focused on:

- a) An evaluation of the role of tropical moisture exports and atmospheric circulation in determining extreme precipitation and floods
- b) A statistical decomposition of SST and SLP fields using correlation networks and its influence on extreme precipitation
- c) An analysis of the space-time inter-annual variability of extreme precipitation (HADEX) and extreme streamflow (USGS HCDN sites)
- d) A comparison of the HADEX extreme precipitation and CMIP5 retrospective and future projections

2. Highlights of Accomplishments:

The project focused on identifying the atmospheric moisture transport and precipitation mechanisms associated with extreme floods in the US and elsewhere and understanding the associated dynamics with a view to prediction of such conditions. This has been the single largest contribution of the project, and we have focused on identifying the role of tropical moisture exports and modeling them for 4 locations. The work establishes the basis for linking climate to synoptic circulation to extreme precipitation and hence to floods.

In the first two papers listed below, we focused on the largest floods in the Ohio River basin and identified that their atmospheric causes were coherent across events, and corresponded to a specific dipole pattern that leads to a wavetrain of moisture and precipitation into the basin every 4 to 7 days in the March-May period. Teleconnections with the Pacific and Atlantic Oceans were identified at different time scales, and lead times. A parallel analysis demonstrated that similar atmospheric circulation features govern the extreme and persistent floods in the UK.

The third paper analyzed a particular extreme flood in France and we were able to identify the tropical moisture exports associated with this flood and other regional precipitation extremes. A synoptic circulation mode that influences such events in the region was identified from a PCA of mid-latitude circulation fields and was shown to provide significant out-of-sample prediction skill for the day by day sequence of extreme precipitation in 1995 that corresponded to the extreme flood event.

The fourth paper establishes the climatology of tropical moisture exports and sources for the North East USA, as well as the associated seasonal circulation mechanisms and their

modulation by known climatic patterns. This work is being extended into a formal Bayesian prediction model.

The seventh paper is most recent, and it pursues an analysis of the tropical moisture exports into N. California and their use for regional prediction of extreme precipitation frequency across 19 sites. Significant skill for out of sample predictions in space and time is demonstrated, provided a basis for how a multi-stage model for climate and precipitation for downscaling climate change or seasonal scenarios could be developed.

In paper 5, we focused on an identification of the potential moisture source teleconnections for extreme precipitation across the world. A correlation network was identified for SST and SLP and its utility in prediction of extreme precipitation over the next 30 days was demonstrated in a cross validated setting.

Finally, in paper 6 we explore the inter-annual and longer frequency structure in extreme precipitation and floods over the continental United States using Wavelet and Clustering methods, and identify groupings of regions that share common space-time-frequency structure in the occurrence of extreme precipitation and floods. These provide a basis for regionalization by putative mechanisms as associated with the inter-annual and decadal structure.

In summary, the project contributed to a mechanistic understanding and prediction of extreme precipitation and floods, using empirical analyses performed on observed and modeled atmospheric circulation patterns.

3. Publications in refereed journals from the Project

1. Robertson, Andrew W. , Yochanan Kushnir, Upmanu Lall, and Jennifer Nakamura, 2013, Weather and Climatic Drivers of Extreme Flooding Events over the Midwest of the United States, AGU Monograph "Extreme Events: Observations, Modeling and Economics", M. Chavez, M. Ghil, J. Urrutia-Fucugauchi (Eds.), sub judice.
2. Nakamura, Jennifer, Upmanu Lall, Yochanan Kushnir, Andrew W. Robertson, Richard Seager, 2013: Dynamical structure of extreme floods in the U.S. midwest and the United Kingdom. *J. Hydrometeor*, 14, 485–504. doi: <http://dx.doi.org/10.1175/JHM-D-12-059.1>
3. Lu, M., Lall, U., Schwartz, A., & Kwon, H. (2013). Precipitation predictability associated with tropical moisture exports and circulation patterns for a major flood in France in 1995. *Water Resources Research*, 49(10), 6381-6392. *This paper was selected as a AGU featured article*
4. Lu, M., and U. Lall, Tropical Moisture Exports, Extreme Precipitation and Floods in Northeast US, in revision for *Water Resources Research*.
5. Lu, M., U. Lall, J. Kawale, S. Liess and V. Kumar, Predicting 30 Days Extreme Precipitation using a Global SST-SLP Correlation Network, *J. of Climate*, *submitted*
6. Lall, U., F. Cioffi, N. Devineni, and Mengqian Lu, Decadal variability in US Floods and Extreme Rainfall, in preparation for submission to *Water Resources Research*.

7. Steinschneider, S., and U. Lall, A hierarchical Bayesian regional model for non-stationary precipitation extremes in Northern California conditioned on tropical moisture exports, submitted to *Water Resources Research*.

Paper 7 is attached – the others were provided in prior reports

4. **PI Contact Information.** Upmanu Lall ula2@columbia.edu, Yochanan Kushnir kushnir@ldeo.columbia.edu, Andrew Robertson awr@iri.columbia.edu, Jennie Nakamura jennie@ldeo.columbia.edu

A hierarchical Bayesian regional model for non-stationary precipitation extremes in Northern California conditioned on tropical moisture exports

Scott Steinschneider¹, Upmanu Lall²

1. Postdoctoral Fellow, 500 West 120th Street, Department of Earth & Environmental Engineering, Columbia University, New York, NY, 10027. Email: scottsteinschneider@gmail.com, Phone: 201-91301964 (Corresponding Author)
2. Professor, 500 West 120th Street, Department of Earth & Environmental Engineering, Columbia University, New York, NY, 10027. Email: ula2@columbia.edu, Phone: 212-854-7081

Abstract

Warm, moist, and longitudinally confined tropical air masses are being linked to some of the most extreme precipitation and flooding events in the mid-latitudes. The inter-annual frequency and intensity of such atmospheric rivers (ARs), or tropical moisture exports (TMEs), are connected to the risk of extreme precipitation events in areas where moisture convergence occurs. This study presents a non-stationary, regional frequency analysis of precipitation extremes in Northern California that is conditioned on the inter-annual variability of TMEs entering the region. Parameters of a multi-site peaks-over-threshold model are allowed to vary conditional on the integrated moisture delivery from TMEs over the Western U.S. Parameters are also related to time-invariant, local characteristics to facilitate regionalization to ungauged sites. The model is developed and calibrated in a hierarchical Bayesian framework to support partial pooling and enhance regionalization skill. The model is cross-validated along with two alternative, increasingly parsimonious formulations to assess the additional skill provided by the covariates. Climate diagnostics are also used to better understand the instances where TMEs fail to explain variations in rainfall extremes to provide a path forward for further model improvement. The modeling structure is designed to link seasonal forecasting and long-term projections of TMEs directly to regional models of extremes used for risk estimation.

1. Introduction

The challenges that extreme hydroclimate events pose to sustainable development around the world are both severe and growing. Recent floods in the United States [Dirmeyer and Kinter, 2010], Pakistan [Houze et al., 2011], China [Zong and Chen, 2000; Lin et al., 2010], Europe [Ulbrich et al., 2003; van den Honert and McAneney, 2011], United Kingdom [Marsh, 2004; Slingo et al., 2014], and Thailand [Komori et al., 2012] have led to an alarming rise in humanitarian crises and societal costs [Munich Re, 2012], underscoring the need to better understand the nonstationary characteristics of these phenomena. The cost-effective design of reliable engineering solutions and financial mechanisms (e.g., insurance products) over annual to decadal planning horizons depends on this scientific understanding.

Nonstationarity in hydroclimate extremes has emerged as a prevailing issue in water resources engineering and hydrology over the past two decades [Olsen et al., 1999; Koutsoyiannis, 2006; Khaliq et al., 2006; Villarini et al., 2009]. Large-scale oscillations in the climate system (e.g., ENSO, PDO) have been shown to modulate the frequency distributions of extreme hydroclimate events around the world, challenging the statistical assumptions underlying classical extreme value frequency analysis [Jain and Lall, 2000; Kwon et al., 2009; Alexander et al., 2009; Lima and Lall, 2010; Ward et al., 2010]. This has important implications for operational water management decisions related to the clustering of extreme events, and also for under/design or over/design of flood control projects if the impact of low frequency climate oscillations is not analyzed [Jain and Lall, 2001]. Projections of future climate change using numerical modeling of the coupled ocean-atmosphere system also suggest that extreme events will worsen in the future due to increased greenhouse gas concentrations [Milly et al., 2002; Hirabayashi et al., 2013]. Support for this argument is often driven by thermodynamic considerations, i.e., an increase in the water holding capacity of the atmosphere [Lenderink, and van Meijgaard, 2008; Muller et al., 2011; Romps, 2011; Berg et al., 2013]. Unfortunately, due to model structural errors, parameterization uncertainty, and the challenges of resolving sub-grid precipitation processes, model integrations tend to exhibit substantial bias in the spatiotemporal distribution of regional precipitation, particularly for tail events, impeding a straightforward inference on future extreme floods due to changes in atmospheric circulation [Karamperidou et al., 2012; Fischer et al., 2013].

When trends are present in the instrumental record, statistical methods for nonstationary frequency analysis are available [Coles, 2001; Khaliq et al., 2006]. However, it is difficult to support the extrapolation of trends for inferences of future recurrence intervals [Koutsoyiannis and Montanari, 2014]. Recently, Merz et al., [2014] argue that nonstationary statistical techniques need to be complemented by an understanding of the causal mechanisms and dominant climatic processes that modulate extreme event risk. Such an understanding is indeed necessary to develop a consistent analytical framework to explain and predict oscillations and trends in extremes using both historic observations and long-term climate projections from atmosphere-ocean general circulation models (AOGCMs). This approach may be particularly relevant when trying to estimate future extreme event risk at ungaged sites that require model regionalization from gaged locations [Lima and Lall, 2010]. Consequently, examining the causal chain of events that leads to extreme floods is important for an understanding of the associated nonstationarity. This is a goal of the current paper.

In the midlatitudes, organized and persistent moisture transport sourced from the tropical oceans has been identified as one climate mechanism that can significantly influence the generation of hydroclimatic extremes. This class of atmospheric phenomena is referred to as tropical moisture exports (TMEs), or atmospheric rivers (ARs) [Zhu and Newell, 1994] when taking the form of planetary-scale filaments of atmospheric moisture. ARs have been estimated to account for almost 90% of global meridional moisture conveyance [Zhu and Newell, 1998] and have been shown to exhibit substantial inter-annual and decadal variability in certain locations [Dettinger and Cayan, 2014], at least in the recent past [St. George and Ault, 2011]. Knippertz and Wernli [2010] established a TME climatology for the

northern hemisphere, identifying four activity centers, including: 1) the Pineapple Express (PE) that connects tropical moisture near Hawaii to the western coast of North America; 2) the West Pacific (WP) TME that influences extreme rainfall in East Asia; 3) a moisture transport maximum over the Great Plains (GP) affecting the central U.S.; and 4) the Gulf Stream (GS) TME located near the western inflow portion of North Atlantic storm tracks. TME climatology was also recently extended to the southern hemisphere [Knippertz et al., 2012].

A number of hydroclimatic analyses have linked ARs and TMEs to large-scale rainfall and flooding events in the Western U.S. (PE) [Neiman et al., 2008; Dettinger et al., 2011; Neiman et al., 2011; Dettinger, 2013; Ralph et al., 2013], Central and Eastern U.S. [Dirmeyer and Kinter, 2009; Dirmeyer and Kinter, 2010; Moore et al., 2012; Nakamura et al., 2012; Lavers and Villarini, 2013; Lackmann, 2013] (GP), United Kingdom [Lavers et al., 2011] (GS), and Western Europe [Lu et al., 2013] (GS). The flood events associated with these features are often the most extreme in the instrumental record and produce the greatest socioeconomic damage. To date, however, the authors are unaware of any work that integrates the temporal behavior of TME dynamics into extreme event frequency modeling of teleconnected midlatitude regions, despite their importance in large-scale flood genesis.

In the context of this research gap, this paper seeks to study the following questions using a case study of precipitation extremes in Northern California:

1. How does annually/seasonally integrated moisture transport related to TMEs influence both the frequency and magnitude of extreme rainfall events over the region?
2. Can the inter-annual variability of an annualized TME index of regional moisture delivery help improve extreme rainfall frequency models at gaged sites and regionalization of these models to ungaged sites?
3. When does such an annual TME index fail to explain variations in extreme rainfall, and do these discrepancies originate from inadequacies in the index or from other prevailing extra-tropical atmospheric circulation during these periods?

These questions are explored in the context of a multi-site peaks-over-threshold (POT) model of precipitation extremes. The primary goal of this work is to establish a direct link between an emerging process-based knowledge of organized, large-scale climate dynamics linked to widespread flooding in the midlatitudes and extreme event frequency modeling often used for risk estimation. The modeling framework is proposed as an approach that can support both seasonal predictions and long-term projections of future extreme event risk based on developing knowledge of these mid-latitude causal mechanisms.

The remainder of the paper is organized as follows. Section 2 outlines the data used in the

study, while methods are presented in Section 3. Section 4 presents the results of the analysis, and the paper concludes in Section 5 with a discussion of potential avenues for future work.

2. Data

The climate data used in this analysis are taken from the 6-hourly ERA-Interim global atmospheric reanalysis dataset with 3-hourly forecasts of surface parameters [Dee et al., 2011]. These data have a spectral T255 horizontal resolution (0.75° , or $\sim 80\text{km}$), with a total of 39 grid cells covering the study domain in Northern California (see Figure 1, inset). Cumulative total precipitation estimates are gathered for times 00:00 UTC and 12:00 UTC at the last 3-hourly forecast time step and are summed to estimate the cumulative precipitation for each day. The data are gathered over the period from November 1, 1979 to March 31, 2013 and are restricted only to the winter months (November-March), as this season contains essentially all extreme precipitation events in Northern California. A total of 34 winter seasons are available. Precipitation from the ERA-Interim reanalysis is chosen over other gridded precipitation products based on station data to be consistent with the TME data which is also derived from the ERA-Interim reanalysis.

For each of the 39 grid cells included in the analysis, a series of time-invariant characteristics are also gathered, including the latitude and longitude of each grid cell, as well as the mean and standard deviation of orography, both measured in units of m^2/s^2 .

Finally, data on TMEs over the region are derived from the global climatology presented in Knippertz et al. [2012], also based on the ERA-Interim reanalysis dataset. The TME identification procedure is presented in detail in Knippertz and Wernli [2010] and is only briefly summarized here. Using the software package Lagrangian Analysis Tool (LAGRANTO) [Wernli and Davies, 1997], trajectories of TMEs are identified from vectors of three-dimensional wind fields. First, one-day forward trajectories are calculated from every $100\text{ km} \times 100\text{ km} \times 30\text{ hPa}$ box between 0°N and 20°N (i.e., the northern tropics) and within the range of 1000 to 490 hPa on a particular day at 00:00 UTC. The vertical range is limited to 490 hPa because approximately 90% of all atmospheric water vapor is contained below this level. If tracks cross the 20°N latitude threshold on day 1, they are followed for the next 6 days. Those tracks that maintain at least a $100\text{ g kg}^{-1}\text{ m s}^{-1}$ water vapor flux beyond 35°N sometime within the 7 day period are classified as a TME track. This classification scheme promotes the retention of fast moving air parcels to ensure that the parcels maintain their tropical characteristics as they move poleward into the subtropics. Tracks are gathered for each day between November 1, 1979 and March 31, 2013 and are identified by their date of birth in the tropics. After the tracks are identified, a variety of meteorological parameter fields are then interpolated to the position of the trajectory at each 6-hourly time step, for a total of 29 time steps within the 7-day period. Of interest in this study, specific humidity is interpolated to enable the tracking of water flux along each TME trajectory from its source in the tropics to its terminus in the extra-tropics.

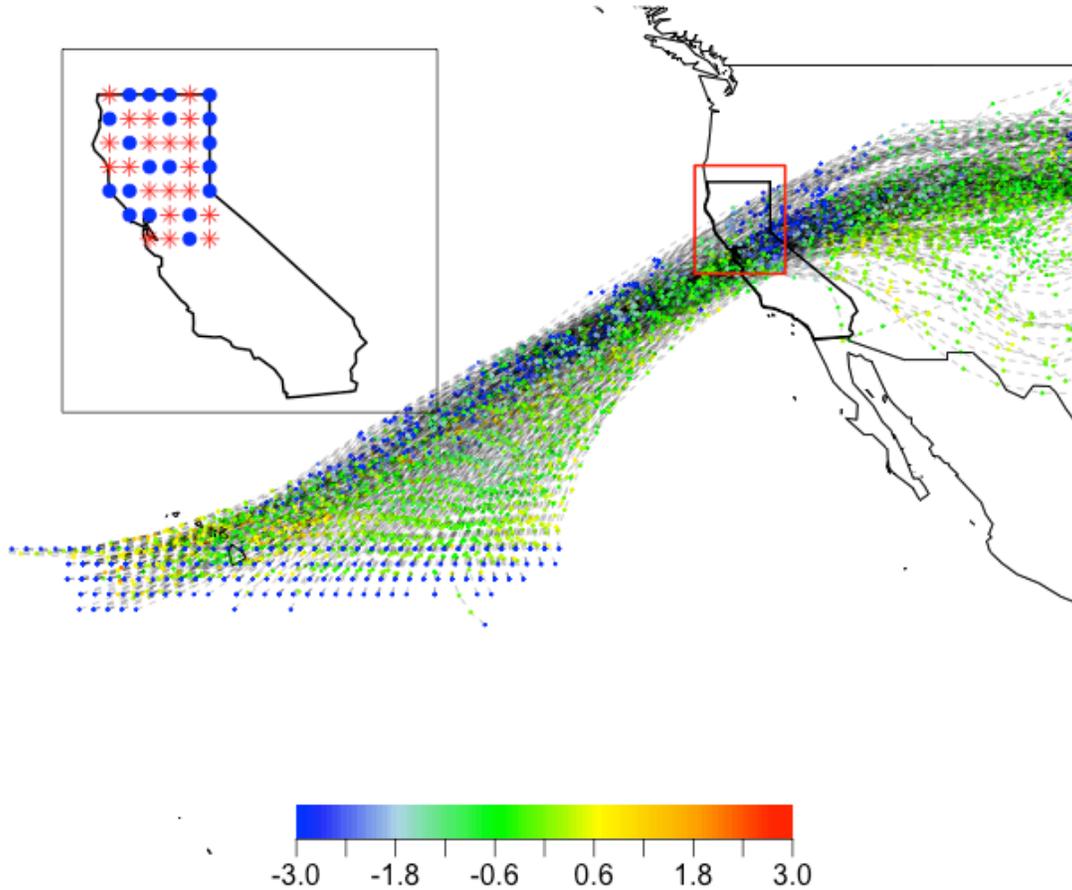


Figure 1. TME tracks passing over California on January 1, 1997. All tracks pass through the domain delimited by the red box. Colors indicate the change in specific humidity along the TME trajectory ($\Delta g \text{ kg}^{-1}$), with negative values indicating a release of water. Inset: The 39 reanalysis grid cells included in the analysis. Blue circles (red asterisks) indicate sites included (excluded) during model fitting.

3. Methods

To address our research questions, we characterize the relationship between the frequency and magnitude of extreme precipitation events in Northern California with an index of annual TME moisture delivery over the region. We develop a regional Peak over Thresholds (POT) model of precipitation extremes that conditions the frequency and magnitude of exceedances on the annual TME index. This model also uses site-specific characteristics to regionalize model parameters to ungaged sites. Two nested, increasingly parsimonious models are also developed for comparison to determine the additional benefit provided by the covariates for both calibration and validation (i.e., “ungaged”) sites. Model results are then evaluated to assess in which years the TME information is a useful contributor or not for assessing extreme precipitation risk. For those years where skill is lowest/highest, an analysis of the dominant features of hemispheric atmospheric circulation and regional precipitation are explored.

3.1. Partial duration series and TME index

To analyze extreme precipitation in Northern California, a POT modeling framework is proposed for the number of occurrences Y_{it} and magnitude X_{it} of precipitation events that exceed some threshold μ_i , where i indexes the precipitation site, and $t=1\dots T$ is an index of each winter season from 1979 to 2013. All analyses are conducted on a random subset of 19 precipitation sites selected from the original set of 39, leaving 20 sites for model validation (see Figure 1 inset). Partial duration series are developed for the 19 calibration sites by first selecting the threshold μ_i , and then calculating the number of threshold exceedances $Y_{i,t}$ in each winter season. The magnitude of each exceedance over μ_i in year t is denoted $X_{i,t}^k = P_{i,t}^k - \mu_i$, where $k=1, \dots, K_{it}$, $Y_{i,t}$ and $P_{i,t}^k$ is the precipitation value on the day of the k^{th} exceedance for year t and site i , and K_{it} is the number of exceedances of the threshold in year t at site i .

To select the threshold μ_i , a series of candidate levels corresponding to precipitation percentiles (90th, 91st, 92nd, ..., 99th) are first calculated and superimposed on mean residual life plots for each calibration site. These plots display the mean excess (i.e., the mean of the precipitation excesses over some threshold μ , less the threshold) against a range of different threshold values. An appropriate threshold is identified as the value beyond which the mean excesses are linear in μ [Coles, 2001]. The mean residual life plots for all sites are visually examined with the percentile values superimposed, and the percentile for which the threshold most consistently passes the linearity criterion for all 19 sites, is selected as the basis for μ_i . Using this method, all sites have approximately the same number of exceedances. To avoid double counting precipitation extremes from the same storm event, precipitation exceedances with less than 3 days of separation are considered a single event, and the maximum exceedance during that time period is retained as the precipitation peak.

An index is created for the total wintertime moisture released from TMEs over Northern California. First, all tracks from the northern hemisphere TME climatology database that pass over a domain between 36.5-43°N and 125-119°W are identified (see red box in Figure 1). For each track, the change in moisture is calculated as $\Delta Q_j = Q_j - Q_{j-1}$, where $j \in \{0, 6, 12, \dots, 168 \text{ hours}\}$ indexes the time along the trajectory when that track is located over the domain and Q is the specific humidity in g/kg. If ΔQ_j is negative, this indicates that moisture was released at time j along the trajectory, while a positive value indicates that moisture was added from evaporation. Given j and the date of birth of a track, each ΔQ_j can be associated with a particular day in the record. All ΔQ values are summed across all tracks for each day of winter between 1979 and 2013 to develop a single, daily time series of ΔQ for the region. The daily time series can then be summed for each winter season to create a single annual index, TME_t . This index represents the cumulative moisture delivery from TMEs averaged over the entire domain, rather than for a particular location, and therefore can be used as a covariate for all POT models across precipitation sites. We note that while the regional TME index likely averages out important spatial variations in TME moisture delivery, which is a clear shortcoming, the application of a single TME covariate in all POT models will likely lead to more robust relationships that are less sensitive to spatial biases in the TME database.

3.2. POT model development

The number of occurrences $Y_{i,t}$ and magnitude $X_{i,t}^k$ of precipitation exceedances are investigated using three nested models (M0, M1, M2) with increasing complexity that accounts for additional information provided by site-specific characteristics and the inter-annual variability of TMEs. The number of peak occurrences $Y_{i,t}$ for a particular site are often assumed to follow a Poisson distribution [Shane and Lynn, 1964], while the magnitude of exceedances $X_{i,t}^k$ are modeled using a generalized Pareto distribution (GPD) or exponential distribution (a GPD with zero shape) [Pickands, 1975]. Hereafter we continue with the simpler exponential model, although the general approach is easily extendable to the GPD. The density functions of the Poisson and exponential distributions are given as:

$$f(Y_{i,t}|\lambda_i, \mathbf{\Lambda}, \mathbf{W}) = \lambda_i(\mathbf{\Lambda}, \mathbf{W})^{Y_{i,t}} \frac{e^{-\lambda_i(\mathbf{\Lambda}, \mathbf{W})}}{Y_{i,t}!} \quad (1)$$

$$f(X_{i,t}^k|\theta_i, \mathbf{\Theta}, \mathbf{W}) = \theta_i(\mathbf{\Theta}, \mathbf{W}) e^{-\theta_i(\mathbf{\Theta}, \mathbf{W})X_{i,t}^k} \quad (2)$$

The Poisson parameters λ_i and exponential parameters θ_i are each assumed to be random variables with distributions conditional on a vector of hyperparameters $\mathbf{\Lambda}$ and $\mathbf{\Theta}$, respectively, as well as covariates \mathbf{W} . The hyperparameters are assigned a prior distribution $P(\mathbf{\Lambda}, \mathbf{\Theta})$, and the joint posterior distribution of all model parameters across sites can then be described using Bayes Theorem:

$$P(\lambda, \theta, \mathbf{\Lambda}, \mathbf{\Theta} | Y, \mathbf{X}, \mathbf{W}) \propto \prod_{i=1}^N \left(P(\lambda_i, \theta_i | \mathbf{\Lambda}, \mathbf{\Theta}, \mathbf{W}) \prod_{t=1}^T \left(f(Y_{i,t} | \lambda_i, \mathbf{\Lambda}, \mathbf{W}) \prod_{k \leq Y_{i,t}} f(X_{i,t}^k | \theta_i, \mathbf{\Theta}, \mathbf{W}) \right) \right) P(\mathbf{\Lambda}, \mathbf{\Theta} | \mathbf{W}) \quad (3)$$

The hierarchical Bayesian structure presented in equation 3 allows partial pooling of information across the N calibration sites by having both θ_i and λ_i distributed according to a regional parent distribution $P(\lambda_i, \theta_i | \mathbf{\Lambda}, \mathbf{\Theta}, \mathbf{W})$. The three models M0, M1, and M2 increasingly allow for additional pooling in the hierarchy by adding structure to $P(\lambda_i, \theta_i | \mathbf{\Lambda}, \mathbf{\Theta}, \mathbf{W})$.

The simplest model, M0, presents a basic hierarchical Bayesian structure for λ_i and θ_i , where the parameters from each site i are assumed to be drawn from lognormal regional parent distributions:

$$\alpha_i \sim N(\mu_\alpha, \tau_\alpha) \quad (4)$$

$$\gamma_i \sim N(\mu_\gamma, \tau_\gamma) \quad (5)$$

$$\lambda_i = \exp(\alpha_i); \theta_i = \exp(\gamma_i) \quad (6)$$

Given the exponential link functions in equation 6, the mean hyperparameters μ_α and μ_γ are given vague uniform distributions between -10 and 10, while the precision hyperparameters τ_α and τ_γ are both given weakly informative gamma distributions with shape and rate equal to 0.001. Here, we assume that the frequency and intensity of extremes across sites are neither identically distributed nor are they completely unrelated

to one another. Instead, the distributions across sites are exchangeable, such that they can vary from site to site with some structure described by the distributions $N(\mu_\alpha, \tau_\alpha)$ and $N(\mu_\gamma, \tau_\gamma)$, but this variation is not predictable.

The intermediate model, M1, adds an additional level of complexity to M0 by allowing the hyperparameters μ_α and μ_γ to vary by site depending on local characteristics of each precipitation grid cell, specifically the latitude, longitude, average elevation, and standard deviation of elevation. These physical characteristics are time-invariant and determine how position relative to storm tracks and orographic lift alter the extreme precipitation characteristics at a site. Since these regressors are highly correlated, their information is summarized using the first R principal components that explain the majority of variance in the original dataset. The row vector z_i of length R contains the scores for site i. The parameters θ_i and λ_i are then allowed to vary depending on these covariates by making the mean values μ_α and μ_γ linear functions of z_i :

$$\mu_{\alpha,i} = \beta_{\mu_\alpha}^0 + z_i \boldsymbol{\beta}_{\mu_\alpha}^1 \quad (7)$$

$$\mu_{\gamma,i} = \beta_{\mu_\gamma}^0 + z_i \boldsymbol{\beta}_{\mu_\gamma}^1 \quad (8)$$

All regression coefficients are assigned vague uniform priors between -10 and 10.

Finally, the most complex model, M2, allows the parameters θ_i and λ_i to vary temporally based on the inter-annual variability of annually integrated TME moisture deliveries over each winter season, in addition to site-specific characteristics:

$$\lambda_{i,t} = \exp(\alpha_i + \beta_{\lambda,i}^{TME} \times TME_t) \quad (9)$$

$$\theta_{i,t} = \exp(\gamma_i + \beta_{\theta,i}^{TME} \times TME_t) \quad (10)$$

Here, the parameters α_i and γ_i follow the same conditional distributions as in M1, but now the Poisson and exponential models can also vary from year to year depending on the TME index through coefficients $\beta_{\lambda,i}^{TME}$ and $\beta_{\theta,i}^{TME}$. The average frequency of extremes increases as the coefficient $\beta_{\lambda,i}^{TME}$ increases, while the converse is true for the magnitude of extremes and $\beta_{\theta,i}^{TME}$. Because different sites may be influenced differently by TMEs, these coefficients are allowed to vary depending on local characteristics:

$$\beta_{\lambda,i}^{TME} \sim N(\pi_{\beta_\lambda}^0 + z_i \boldsymbol{\pi}_{\beta_\lambda}^1, \tau_{\beta_\lambda}^{TME}) \quad (11)$$

$$\beta_{\theta,i}^{TME} \sim N(\pi_{\beta_\theta}^0 + z_i \boldsymbol{\pi}_{\beta_\theta}^1, \tau_{\beta_\theta}^{TME}) \quad (12)$$

Here, $\pi_{\beta_\lambda}^0, \pi_{\beta_\theta}^0$ are scalars and $\boldsymbol{\pi}_{\beta_\lambda}^1, \boldsymbol{\pi}_{\beta_\theta}^1$ are column vectors of length R, all with uniform priors between -10 and 10, while the precision parameters $\tau_{\beta_\lambda}^{TME}, \tau_{\beta_\theta}^{TME}$ are again given weakly informative gamma distributions with shape and rate equal to 0.001. Equations 9-12 allow the Poisson and exponential parameters for each site to vary through time as a function of TMEs, but this relationship is assumed to be different across sites and this difference is predictable based on the properties of each site. Table 1 provides a

summary of the key similarities and differences of the three nested models, M0, M1, and M2.

The fitted POT model in M2 can be used for non-stationary extreme precipitation probability estimation. The q -year exceedance event, $P_{i,t}^q$, for the i^{th} site in year t can be calculated as [Davison and Smith, 1990]:

$$P_{i,t}^q = F_{exp}^{-1} \left(\left(1 - \frac{1}{(\lambda_{i,t}(z_i, TME_t) \times q)} \right) \middle| \theta_{i,t}(z_i, TME_t) \right) \quad (13)$$

where $\lambda_{i,t}(z_i, TME_t)$ and $\theta_{i,t}(z_i, TME_t)$ are the Poisson and exponential parameters conditional on covariates, as appropriate, and F_{exp}^{-1} is the quantile function of the exponential distribution.

For each of M0, M1, and M2, the POT model can be regionalized to unaged sites after being fit to calibration sites using posterior sampling. For M0, the joint distribution of $(\lambda_{i^*}, \theta_{i^*})$ for a new site i^* can be simulated by drawing samples of $\{\mu_\alpha, \mu_\gamma, \tau_\alpha, \tau_\gamma\}$ from their joint posterior distribution, and then sampling values of α_{i^*} and γ_{i^*} from the normal distributions in equations 4 and 5 and transforming to λ_{i^*} and θ_{i^*} with the exponential link function. Posterior predictive samples of occurrences Y_{i^*,t^*} and magnitude X_{i^*,t^*} for a year t^* can then be drawn from the Poisson and exponential models using the simulations of λ_{i^*} and θ_{i^*} . A similar process can be followed for M1 and M2 using posterior samples from the full set of hyperparameters conditional on particular values for the predictors z_{i^*} and TME_{t^*} , as appropriate. For all models, the threshold μ_{i^*} is estimated at a new location through a regression of threshold values at calibration sites on PC scores.

Table 1. Key similarities and differences between the three nested models M0, M1, and M2.

		M0	M1	M2
--	--	----	----	----

Poisson	Time-invariant component	$\alpha_i \sim N(\mu_\alpha, \tau_\alpha)$ μ_α constant across sites	$\alpha_i \sim N(\mu_{\alpha,i}, \tau_\alpha)$ $\mu_{\alpha,i} = f(\mathbf{z}_i)$	$\alpha_i \sim N(\mu_{\alpha,i}, \tau_\alpha)$ $\mu_{\alpha,i} = f(\mathbf{z}_i)$
	Time-variant component	$\beta_{\lambda,i}^{TME} = 0$	$\beta_{\lambda,i}^{TME} = 0$	$\beta_{\lambda,i}^{TME} \sim N(f(\mathbf{z}_i), \tau_{\beta_{\lambda,i}^{TME}})$
	Sampling distribution of data	$\lambda_{i,t} = \exp(\alpha_i + \beta_{\lambda,i}^{TME} \times TME_t)$ $Y_{i,t} \sim Possion(\lambda_{i,t})$		
Exponential	Time-invariant component	$\gamma_i \sim N(\mu_\gamma, \tau_\gamma)$ μ_γ constant across sites	$\gamma_i \sim N(\mu_{\gamma,i}, \tau_\gamma)$ $\mu_{\gamma,i} = f(\mathbf{z}_i)$	$\gamma_i \sim N(\mu_{\gamma,i}, \tau_\gamma)$ $\mu_{\gamma,i} = f(\mathbf{z}_i)$
	Time-variant component	$\beta_{\theta,i}^{TME} = 0$	$\beta_{\theta,i}^{TME} = 0$	$\beta_{\theta,i}^{TME} \sim N(f(\mathbf{z}_i), \tau_{\beta_{\theta,i}^{TME}})$
	Sampling distribution of data	$\theta_{i,t} = \exp(\gamma_i + \beta_{\theta,i}^{TME} \times TME_t)$ $X_{i,t}^k \sim Possion(\theta_{i,t}), k = 1, \dots, Y_{i,t}$		

3.3. Model fitting and evaluation

All of the Bayesian models described above are coded in the STAN probabilistic coding language for full Bayesian inference. Posterior distributions are evaluated using the Hamiltonian Monte Carlo (HMC) sampling method [Duane et al., 1987]. The HMC method utilizes efficient jumping rules that help to suppress local random walk behavior that can slow convergence. Four chains are run for all parameters with over-dispersed initial values using 5,000 burn-in simulations and 5,000 iterations afterwards. Convergence is assessed based on chain mixing using the Gelman and Rubin convergence criterion [Gelman and Rubin, 1992].

Models M0, M1, and M2 are all fit in a leave-one-out cross-validation framework where one year is left out of the analysis and the model is refit using the remaining years of data. The fitted model is then used to forecast the distribution of extreme value data for the excluded year. In this way, the models can be cross-validated not only against the gages that were completely left out of the model fitting process, but also against those sites used for calibration. Furthermore, the leave-one-out framework is applied when testing against the validation gages, so that the fitted models have the dual challenge of predicting the distribution of extremes for both a year and location not included in the fitting process.

Two sets of statistical tests are used to compare the probabilistic estimates provided by models M0, M1, and M2 under cross-validation. The first approach is based on a ranked probabilistic skill score (RPSS), derived from the ranked probability score (RPS). In the discrete case, the RPS is the squared deviation between the cumulative distribution function (CDF) of a probabilistic forecast and the CDF of the corresponding observation over a preselected number of forecast categories [Weigel et al., 2007]. The RPS for the Poisson distribution fit in each year is calculated using J=6 categories for the number of extreme precipitation events in a given year ($j = 0,1,2,3,4,>5$):

$$RPS_{i,t^*}^{s,M} = \sum_{j=1}^J \left(F_f(j|\lambda_{i,-t^*}^{s,M}) - F_o(j|Y_{i,t^*}) \right)^2 \quad (14)$$

Here, F_f is the forecasted Poisson CDF under model M, F_o is the observed CDF, which is equivalent to a step function from 0 to 1 at the observed value Y_{i,t^*} , t^* is the excluded year in the leave-one-out framework, and $-t^*$ represents the remaining years used during the model fit. The parameter $\lambda_{i,-t^*}^{s,M}$ is a simulated value of the Poisson parameter for site i from its posterior distribution fit without data from year t^* under model M, with $s = 1, \dots, S$ denoting the number of simulations. The RPSS can then be used to compare the RPS for any two models, for instance, M1 and M2:

$$RPSS_i^{M1,M2} = 1 - \frac{\sum_{t^*=1}^T \frac{1}{S} \sum_{s=1}^S RPS_{i,t^*}^{s,M2}}{\sum_{t^*=1}^T \frac{1}{S} \sum_{s=1}^S RPS_{i,t^*}^{s,M1}} \quad (15)$$

Here, the RPSS between models M1 and M2 for the i^{th} site averages over both the posterior simulations of the RPS for each model and all years of data. If $RPSS_i^{M1,M2} > 0$, this indicates that on average, M2 outperforms M1, while the converse is true for a value less than 0.

When the probabilistic forecast is continuous rather than discrete, as for the exponential model, the RPSS can still be calculated for two models by deriving a continuous form of the RPS (i.e., the CRPS) [Hersbach, 2000]. For the exponential model, the CRPS is given as [Friederichs and Thorarinsdottir, 2012]:

$$CRPS_{i,t^*,k}^{s,M} = X_{i,t^*,k}^k - \frac{1}{\theta_{i,-t^*}^{s,M}} \left(2F_f(X_{i,t^*,k}^k | \theta_{i,-t^*}^{s,M}) - \frac{1}{2} \right) \quad (16)$$

Here, $X_{i,t^*,k}^k$ is the k^{th} exceedance peak in year t^* for site i, F_f is the forecasted exponential CDF under model M, and $\theta_{i,-t^*}^{s,M}$ is a simulated value of the exponential parameter for site i from its posterior distribution. The continuous RPSS (CRPSS) can then be calculated for any two models according to equation 15.

In addition to the CRPSS, a probability integral transform procedure [Laio and Tamea, 2007] is used to further evaluate the adequacy of the exponential model fit. In brief, the procedure tests whether probabilistic predictions for a continuous random variable are adequate in a statistical sense. To conduct the test, the cumulative distribution function of the estimated exponential model for year t^* under model M is evaluated with respect to the

observed exceedance magnitudes, $v_{i,t^*}^{k,s,M} = P(X_{i,t^*}^k | \theta_{i,-t^*}^{s,M})$. On a posterior mean basis, the probabilistic predictions are suitable for site i under model M if the $v_{i,t^*}^{k,s,M}$ values averaged across posterior simulations $s \in S$ are distributed uniformly between 0 and 1 across all years. Uniformity is assessed using a Kolmogorov-Smirnov test.

3.4. Diagnostics on model error

For those years when other causal mechanisms besides TMEs are responsible for extreme precipitation events in Northern California, the M2 model may incorrectly underestimate risk. Conversely, M2 may overestimate risk if TME tracks release small or moderate amounts of moisture over the region on many days in the region, thus increasing the TME index, but corresponding precipitation extremes do not occur. To better diagnose these issues and provide an avenue for further model improvement, atmospheric circulation and the distribution of daily ΔQ values are examined during times of poor model performance. To identify those years where M2 is least accurate, the posterior means of $RPS_{i,t^*}^{s,M2}$ and $CRPS_{i,t^*,k}^{s,M2}$ across all simulations $s \in S$ are calculated for each year and site. Both RPS and CRPS values are then averaged across sites for each year, producing two regional, annual indices that measure M2 performance for the occurrence and magnitude of extremes, respectively. Similar annual indices are also produced for M1. The ratio of the frequency index from M2 and M1 is then calculated for each year. Larger ratios in a given year indicate that M1 predicted the distribution of extreme event occurrences better than M2 for that year, while smaller values suggest the opposite. We record the years in which the 6 largest values occur, i.e., the worst M2 performance compared to M1. A similar analysis is conducted for the region magnitude index from M2 and M1. These years are split into two sets based on whether the frequency distribution was over-predicted or under-predicted, which is determined by comparing the average Poisson and exponential parameters across sites estimated under M1 and M2 for each year. For years of under-prediction, 500 mb wind and geopotential height anomalies over the eastern Pacific and Western United States are composited for dates with extreme rainfall at a minimum of 10 of the 19 calibration sites. These composites will indicate the dominant modes of circulation besides TMEs that drive precipitation extremes. For comparison, the same atmospheric fields are also composited for the dates of major Pineapple Express events in 1982, 1986, 1995, 1997, and 2006, which are all years known to be associated with TME-related floods in California.

For years in which the frequency and magnitude distributions are over-estimated by M2, we examine the distribution of daily ΔQ values for all TME tracks. These results are again contrasted against similar distributions from years with large TME-related extreme events. By comparing these two sets of distributions, the results of this analysis will help determine whether the annual TME index over-simplifies certain aspects of daily TME behavior and consequently misrepresents the link between extreme precipitation events and daily ΔQ values.

4. Results

4.1. Partial duration series and TME index

Mean residual life plots for 4 calibration sites are shown in Figure 2, along with several precipitation percentiles superimposed. The residual life plots of the remaining 15 calibration sites are similar. For almost all sites, mean excesses are linear in μ_i if the threshold is set to the 98th percentile of daily precipitation. This threshold leads to between 77 and 88 exceedance events over the 34 years of record at each site. The final number of exceedances fluctuates by site depending on the extent to which events were declustered for that site. With the threshold selected, the number of exceedances $Y_{i,t}$ and their magnitudes $X_{i,t}^k$ are calculated across all sites. An annual, regional index is then developed for the number of occurrences by averaging $Y_{i,t}$ across all sites for each year. A similar index is created for the magnitude of exceedances by first normalizing $X_{i,t}^k$ for each site and then averaging across sites (Figure 3).

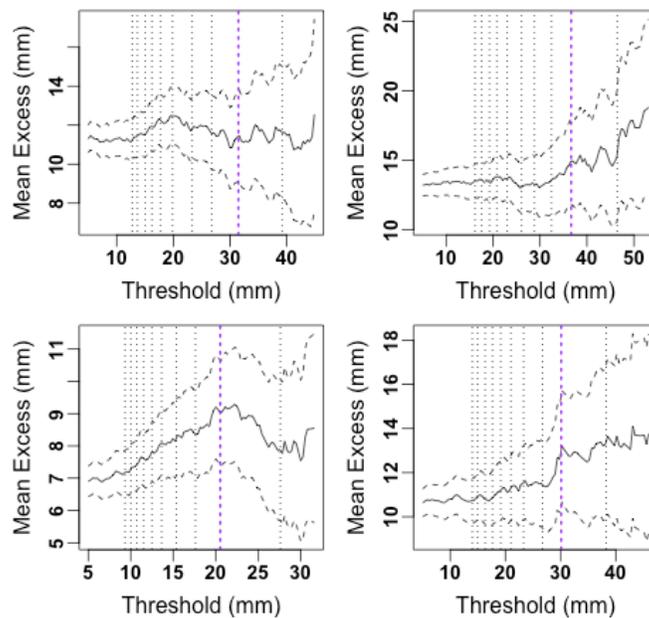


Figure 2. Mean residual life plots for 4 calibration sites. Several precipitation percentiles are superimposed as vertical lines, with the 98th percentile highlighted in bold.

An example of TME tracks and their moisture change for the January 1st storm of 1997 was shown earlier in Figure 1. The TME tracks all originate from the Pacific Ocean near Hawaii, characteristic of a Pineapple Express event. To summarize the activity of these TME events over a season, cumulative moisture deliveries from TME tracks for each winter are integrated into a single, annual index, which is shown along with regional indices for extreme rainfall occurrence and magnitude in Figure 3. The annual TME index is significantly correlated at the 0.01 level with both the regional occurrence (Pearson r of 0.76) and magnitude (Pearson r of 0.49) indices. We note that all indices are relatively high during the winters of 1982, 1986, 1995, 1997, and 2006, which are all years during which California experienced notable floods. However, there are several years in which the number (1985, 1993, 1999, 2003) or magnitude (1987, 1990, 2005, 2013) of exceedances

is large but no corresponding TME peak exists. These discrepancies are discussed further in Section 4.4.

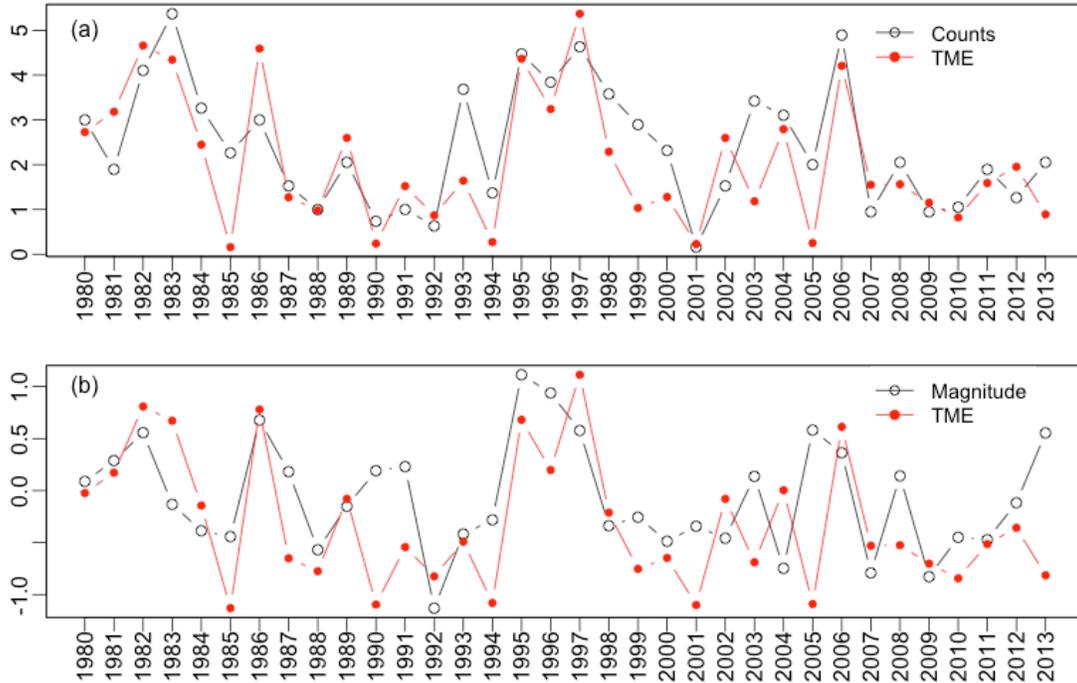


Figure 3. Regional indices of the number (a) and magnitude (b) of extreme rainfall events. The regional TME index is also shown.

Before proceeding with the regional POT model fit, a principal component analysis of site-specific characteristics (latitude, longitude, mean elevation, and standard deviation of elevation) is conducted across the calibration sites. Given the topographic orientation of the Sierra Nevada mountains, the latitude and longitude are both highly correlated with mean elevation (Pearson $r > 0.65$), while the mean and standard deviation of elevation are also moderately correlated (Pearson $r = 0.33$). The first 3 principal components are retained for further analysis, as they account for 96% of the variance (50%, 25%, and 21% respectively).

4.2. Posterior distributions for M2

This section focuses on posterior distributions of key parameters in model M2. Since M0 and M1 are nested sub-models of M2, the results from M2 summarize the general patterns than can be found in the posterior parameter sets across all three models. An inter-model comparison is presented in section 4.3

We first evaluate the posterior estimates of $\beta_{\lambda,i}^{TME}$ and $\beta_{\theta,i}^{TME}$ from M2 that relate the Poisson and exponential models to the TME index (Figure 4). The results are shown for only one fit in the leave-one-out cross-validation. Recall that positive (negative) values of $\beta_{\lambda,i}^{TME}$ ($\beta_{\theta,i}^{TME}$) indicate an increase in the frequency (magnitude) of exceedances. The 90% credible

intervals at all sites for $\beta_{\lambda,i}^{TME}$ and all but one site for $\beta_{\theta,i}^{TME}$ exclude zero, indicating a statistically significant, positive relationship between regional TMEs and both the frequency and magnitude of extremes across the entire domain. For the northeastern site where the 90% credible interval of $\beta_{\theta,i}^{TME}$ includes zero, the 88% credible interval does not. The absolute magnitudes of $\beta_{\lambda,i}^{TME}$ and $\beta_{\theta,i}^{TME}$ exhibit similar spatial gradients across the domain, with larger (smaller) values in the northeast (southwest).

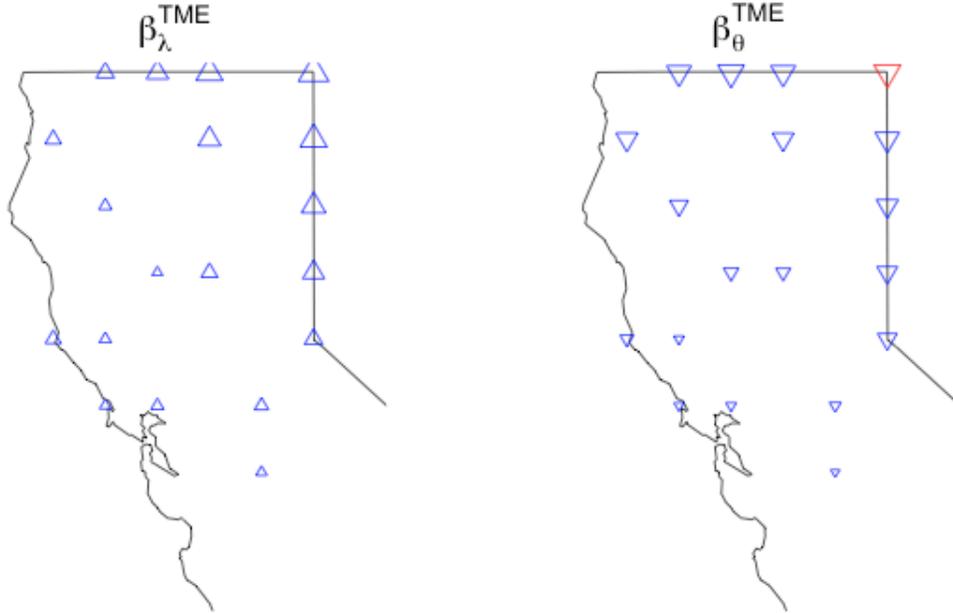


Figure 4. Posterior estimates of $\beta_{\lambda,i}^{TME}$ and $\beta_{\theta,i}^{TME}$ across calibration sites for one cross-validation fit (1980 excluded) of M2. The direction of the triangle indicates the sign of the posterior mean for the coefficient, while the size represents its absolute magnitude. Triangles are colored blue if the 90% credible interval excludes zero.

The spatial gradients for both $\beta_{\lambda,i}^{TME}$ and $\beta_{\theta,i}^{TME}$ suggest that the link between TMEs and the frequency and magnitude of extremes varies systematically across the domain. This is confirmed in Figure 5, which shows the median, 5th, and 95th percentiles of $\pi_{\beta_{\lambda}^{TME}}^1$ and $\pi_{\beta_{\theta}^{TME}}^1$ for each of the cross-validation years. These parameters demonstrate the extent to which $\beta_{\lambda,i}^{TME}$ and $\beta_{\theta,i}^{TME}$ vary with the PCs. For PC1, the 90% credible interval for $\beta_{\theta,i}^{TME}$ almost always excludes zero for all cross-validation years, and the same confidence interval for $\beta_{\lambda,i}^{TME}$ shows a similar, albeit less significant, pattern. No significant relationships emerge for PC2 and PC3. Given the loadings on PC1, the relationships in Figures 5a,d suggest that precipitation extremes at inland, high-elevation areas tend to be more sensitive to TME activity. This is consistent with the notion that orographic lift is needed to initiate precipitation from TMEs [Dettinger, 2011].

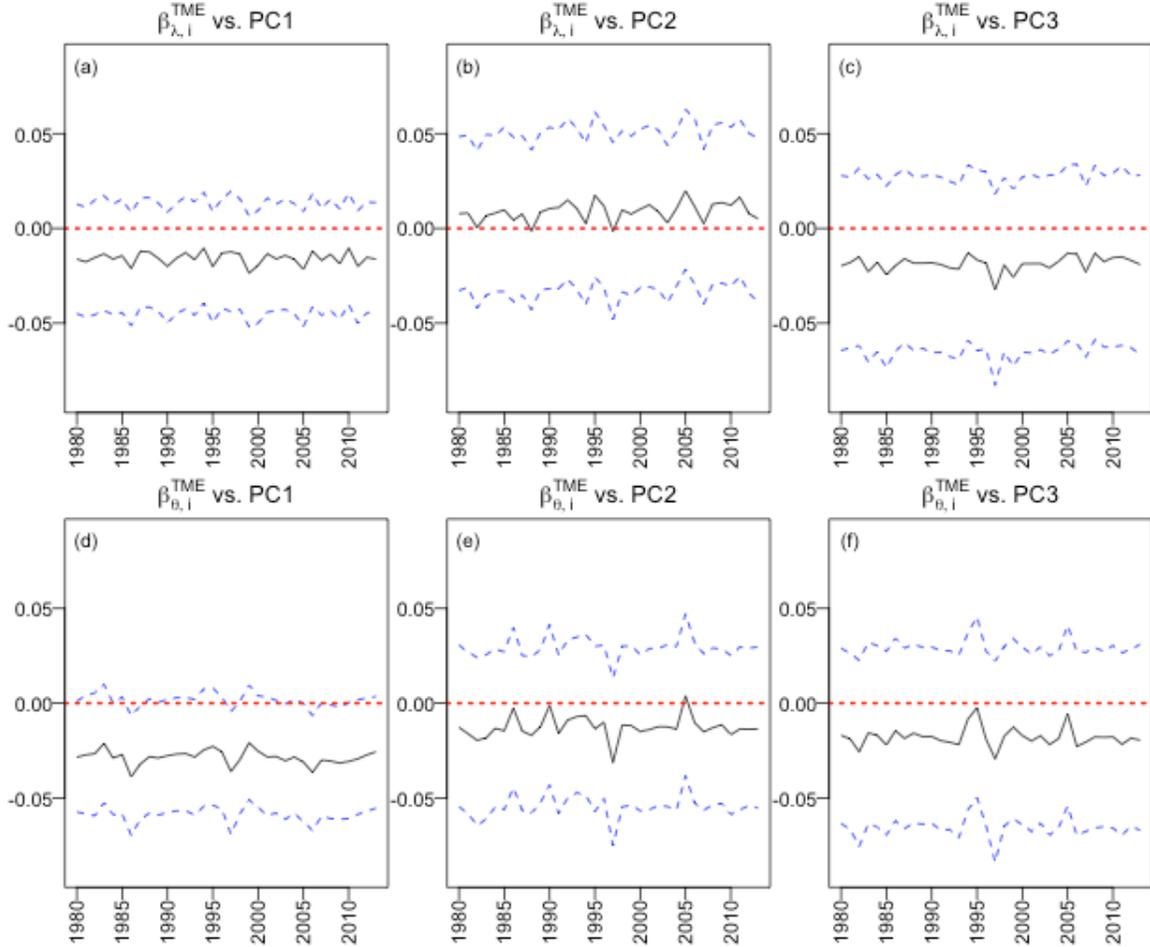


Figure 5. The median (black) and 90% credible interval (dashed blue) of posterior distributions for a-c) $\pi_{\beta_{\lambda}}^1$ and d-f) $\pi_{\beta_{\theta}}^1$ for the fit of M2 in each cross-validation year. A horizontal, dashed red line is shown at zero.

The spatial distribution of posterior estimates for $\mu_{\alpha,i}$ and $\mu_{\gamma,i}$ is shown in Figure 6. These parameters represent how the time-invariant component of Poisson and exponential parameters varies across space. Both parameters tend to exhibit a dipole pattern, with larger absolute magnitudes in the southwest and northeast and smaller magnitudes in the center of the domain, although this pattern is more apparent for $\mu_{\alpha,i}$. Figure 7d-f shows that the posterior distribution of $\beta_{\mu_{\gamma}}^1$, which relates $\mu_{\gamma,i}$ to the PCs across sites, is significantly different from zero for all PCs and cross-validation fits. A much less consistent pattern is found for $\mu_{\alpha,i}$. This parameter shows no real significant variation with any of the PCs (Figures 7a-c). Further analysis (not shown) indicates a similar relationship under M1.

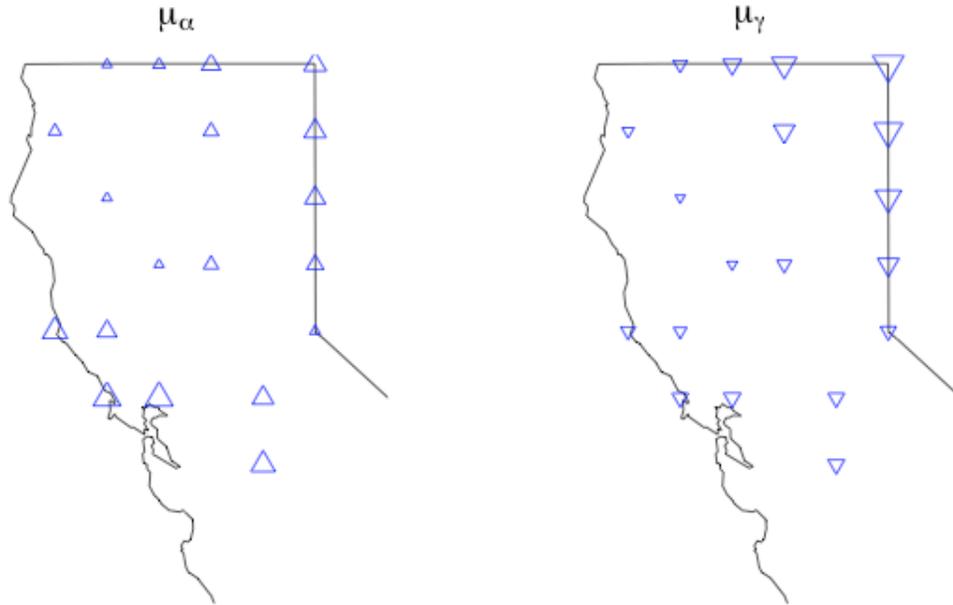


Figure 6. Posterior estimates of $\mu_{\alpha,i}$ and $\mu_{\gamma,i}$ across calibration sites for one cross-validation fit (1980 excluded) of M2. The direction of the triangle indicates the sign of the posterior mean, while the size represents its absolute magnitude. Triangles are colored blue if the 90% credible interval excludes zero.

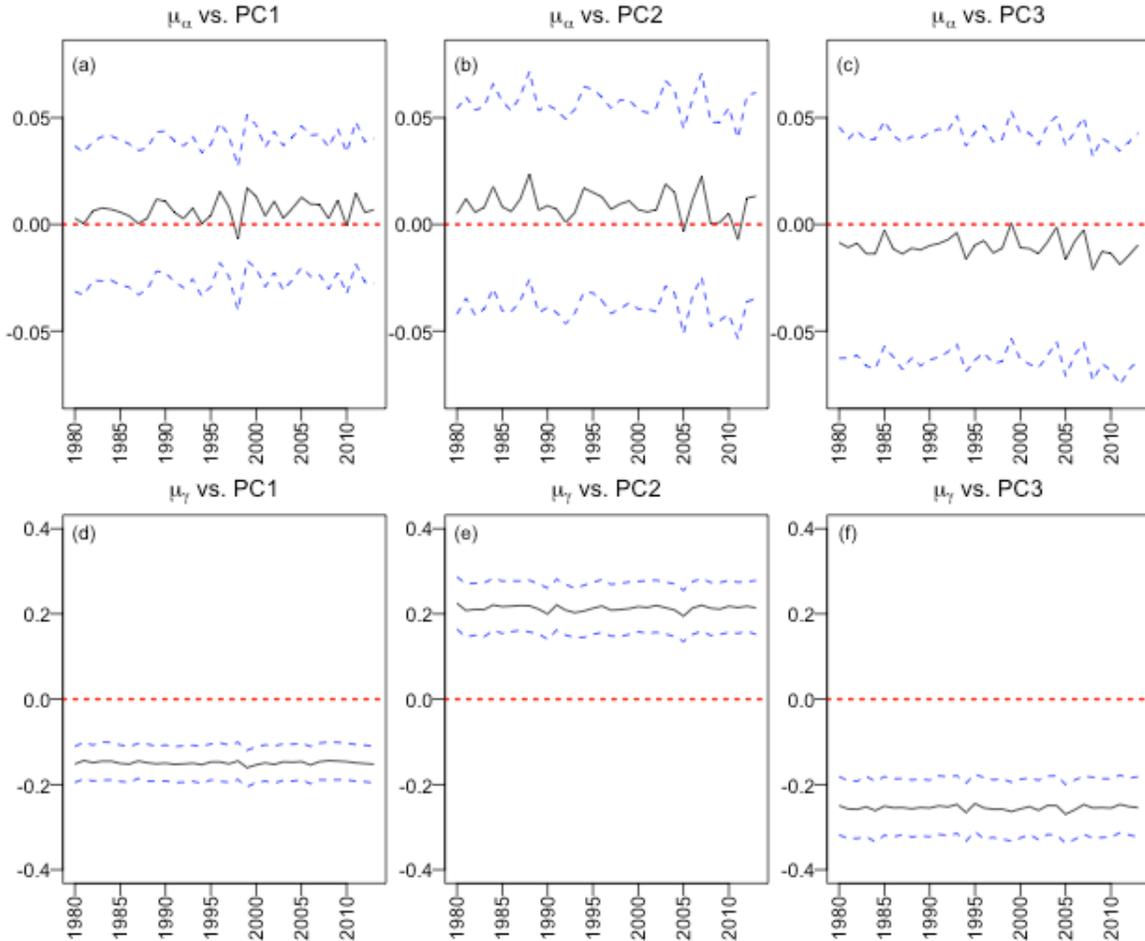


Figure 7. The median (black) and 90% credible interval (dashed blue) of posterior distributions for a) $\beta_{\mu_\alpha}^1$ and b) $\beta_{\mu_\gamma}^1$ for the fit of M2 in each cross-validation year. A horizontal, dashed red line is shown at zero.

The observed decadal variability in the TME index leads to decadal variability in the risk of extremes in Northern California over the historical period. Figure 8 shows the posterior mean and 95% credible interval for the 100-year storm at one site over the past 34 years, as conditioned on the TME index under M2. Also shown is a stationary estimate of the 100-year event, $P_{i,0}^{100}$ for that site based on a maximum likelihood fit of the POT model using all years of data. It is clear that under the assumptions of M2, the risk of extreme rainfall varies substantially from year to year in the observed record as compared to a stationary estimate of extremes. In 1997, the 100-year rainfall event nearly doubles between the posterior mean under M2 and the stationary estimate.

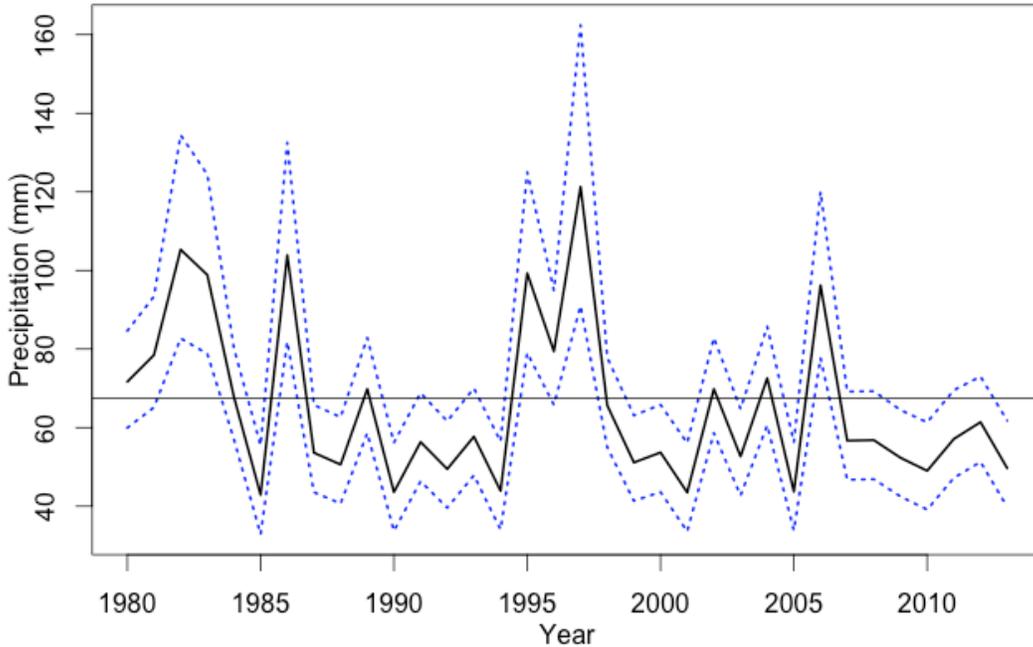


Figure 8. The posterior mean (black) and 95% credible interval (dashed blue) of the 100-year storm estimate under M2. The horizontal line shows a stationary estimate of the 100-year event.

4.3. Inter-model comparison between M0, M1, and M2

The RPSS and CRPSS for the Poisson and exponential cross validation fits are shown in Figure 9. Here, a more complex model (e.g., M2) is always compared against a simpler sub-model (e.g., M1 or M0). A positive RPSS or CRPSS indicates that the more complex model performs favorably for out-of-sample predictions at a particular site. Comparisons between M2 and M1 isolate the additional benefit provided by TME information, while the comparison between M1 and M0 isolates the additional skill provided by including site-specific information into a time-invariant fit of λ_i and θ_i . The comparison between M2 and M0 reveals the benefit provided by considering all covariates.

Several results emerge from Figure 9. First, for both Poisson and exponential fits, M2 performs the best under cross-validation at both calibration and validation sites. This outperformance is most prevalent for exceedance occurrences. On average, M2 outperforms M1 by 20% (14%) at calibration (validation) sites, and this outperformance is consistent across all sites and reaches higher than 25% for a few locations. When considering exceedance magnitudes, the outperformance is much less, though still apparent at verification sites. Here, the average outperformance is 0% (1%) at calibration (validation) sites, with the highest outperformance reaching only 4%. Also, M1 performs better at 7 out of 19 calibration sites, although not by much. However, M2 does show improved performance at 14 out of the 20 verification sites. These results suggest that the addition of TME-based information improves both aspects of the regional POT model, especially for the frequency of peak occurrences.

When comparing M1 and M0, a slight decline in performance is seen for the Poisson model fit under M1. This is somewhat expected given that the Poisson parameter did not ultimately vary substantially with any covariates, yet M1 still propagates the uncertainty in those regression parameters forward (see Figure 7a-c). However, there is significant improvement in the probabilistic forecasts under M1 for exceedance magnitudes, particularly for validation sites not included in the model fit. Here, the additional information provided by the PCs improved the fitted exponential model on average by about 9%, although this value is skewed upward by improvements of 22% and 38% for two validation sites in the northeast of the domain.

Finally, the comparison of the most complex model (M2) with the baseline model (M0) shows strong improvement across both Poisson and exponential fits, particularly for the validation sites. This suggests that a spatiotemporal regionalization of POT models is substantially improved using both large-scale climate phenomena and local-scale characteristics.

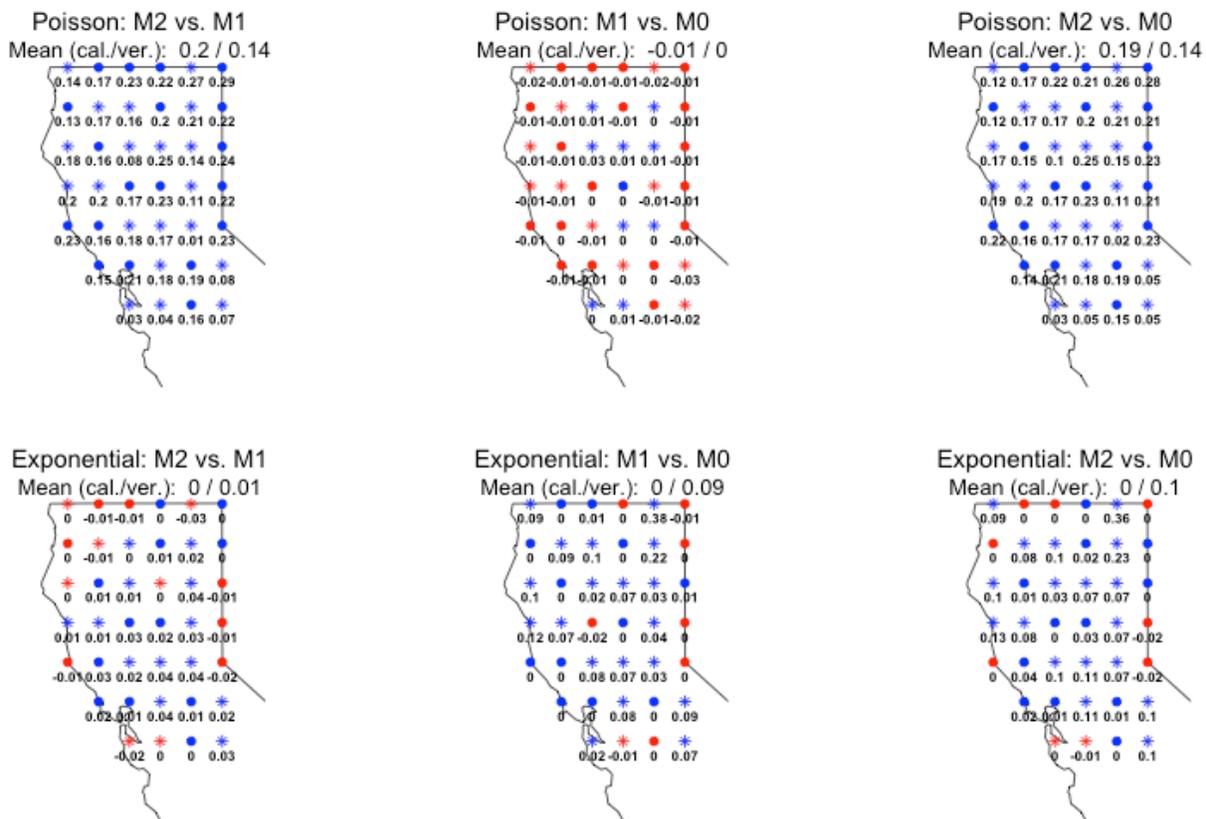


Figure 9. RPSS for the Poisson distribution fit (first row) and CRPSS for the exponential model fit (second row) for different model comparisons. The scores are shown and labeled for calibration (circles) and validation (asterisks) sites. Positive scores (blue) favor the more complicated model, while negative scores (red) suggest the additional complexity

degrades out-of-sample prediction. The mean score across calibration and validation sites is provided in the header.

Since the fit of exponential distributions across the 3 models is more similar than for the Poisson distributions, a secondary evaluation is useful. The results of the PIT analysis on the exponential fit show that all three models have no calibration sites with a p-value less than 0.05 for the Kolmogorov-Smirnov test of uniformity, but this threshold is crossed for 10, 5, and 4 validation sites under models M0, M1, and M2, respectively. For a p-value threshold of 0.10, the number of violations under M0, M1, and M2 changes to 1, 1, and 0 for the calibration sites and 14, 6, and 5 for the validation sites. These results are consistent with those from Figure 8 and suggest that there is a moderate improvement for the exponential model fit between M1 and M0, and a minor improvement between M2 and M1.

Theseresults indicate that model M2 provides the best representation of POT processes across the region. To understand how well this model could perform with more information, M2 is refit using all 39 sites ($M2_{all}$) and compared to the fit based only on the calibration sites ($M2_{regionalized}$). This comparison is shown in Figure 10. We note that it is possible for $M2_{regionalized}$ to slightly outperform $M2_{all}$ at certain sites because partial pooling influences parameters in both models. For the Poisson model, $M2_{all}$ performs substantially better than $M2_{regionalized}$ at most validation locations, with an average outperformance of 20%. For one site, the improvement reaches as high as 60%. Despite the large differences in performance between $M2_{all}$ and $M2_{regionalized}$ at validation sites, the estimated coefficients for TME influence ($\beta_{\lambda,i}^{TME}$) at those sites are extremely similar (Pearson r of 0.99). Rather, the substantial outperformance of $M2_{all}$ reflects the spatial variability of the time-invariant component of the Poisson model ($\mu_{\alpha,i}$) and how poorly it regionalizes across sites in $M2_{regionalized}$ based on the PCs (recall Figure 7a-c). The additional spatial information in $M2_{all}$ is responsible for the higher precision of the associated model coefficients. Conversely, the time-invariant component of the exponential model is much less variable across sites and relates well to all of the PCs, enabling reasonable estimates of $\mu_{\gamma,i}$ across validation sites under $M2_{regionalized}$. This is reflected in the similar performance between $M2_{all}$ and $M2_{regionalized}$ for the exponential fit, with an average outperformance of only 3% at validation sites. We do note, however, that $M2_{all}$ shows a >20% improvement over $M2_{regionalized}$ at two validation sites in the southwest of the domain. Overall, the results suggest that regionalization of POT models across Northern California can be substantially improved with a better understanding of how site-specific characteristics control the time-invariant characteristics of the Poisson model.

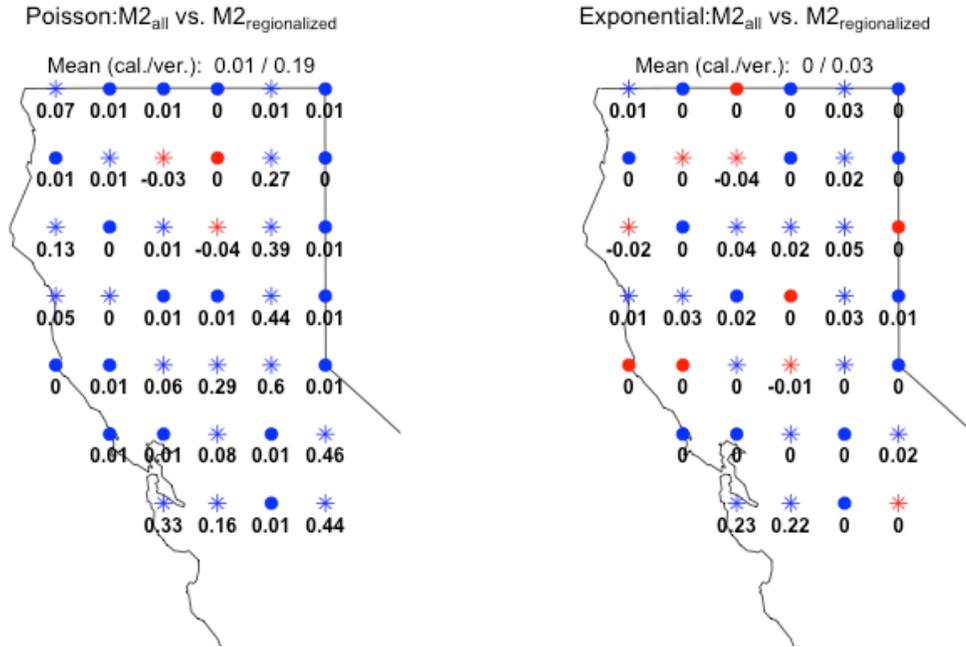


Figure 10. RPSS for the Poisson distribution fit (left) and CRPSS for the exponential model fit (right) for M2 calibrated with all 39 sites ($M2_{all}$) and only the calibration sites ($M2_{regionalized}$). The scores are shown and labeled for calibration (circles) and validation (asterisks) sites. Positive scores (blue) favor $M2_{all}$. The mean score across calibration and validation sites is provided in the header.

4.4. Model error diagnostics

The ratio of posterior mean rank probability scores between M2 and M1 indicates that M2 underperforms M1 by the largest margin in 1981, 1985, 1993, 1999, 2000, and 2003 for the Poisson model fit. For the exponential fit, M2 most underperforms M1 in 1983, 1987, 1990, 2003, 2005, and 2013. Of the poorest fits for the Poisson model, over-prediction only occurs in 1981, while all other years are under-predicted. Similarly, for the exponential model, M2 only over-predicts in 1983 and under-predicts in the remaining years.

Figure 11 shows anomaly composites of atmospheric circulation for extreme event days during years of under-prediction for both the Poisson and exponential models. Similar anomaly composites are also shown for the dates of major PE-related extreme precipitation events. Atmospheric circulation is similar during extreme events in years where the Poisson and exponential models of M2 most underperform M1. The circulation is dominated by an anomalous low off of the northwestern coast of the United States, with a longitudinally extended high to the south. The offshore cyclone drives moisture transport over Northern California from the west and southwest. This pattern is in contrast with circulation during the largest TME events, where the anomalous high to the south extends further over the continental U.S., as does the trough. Moisture is funneled from farther southwest over Hawaii directly to Northern California, forming a characteristic AR event. Figure 11 clearly shows that the POT model conditioned only on TME events is vulnerable

to underestimation when extremes are caused by midlatitude cyclones in the secondary circulation. The model could likely be improved if an additional index was included to represent the variability of frontal systems, although this analysis was not attempted here.

Figure 12 displays the distribution of daily ΔQ values for all TME tracks during years of over-prediction (1981 and 1983), as well as other years containing TME-related extreme events. Interesting, even though 1981 and 1983 have an annual TME index value similar to that found in the other years, they both exhibit the smallest tail behavior of daily ΔQ values. This suggests that in both 1981 and 1983, there were many days with less intense TME tracks that in aggregate delivered substantial amounts of moisture to the region. We also note that 1981 and 1983 have the smallest ratio of the annual TME index value to the number of days with non-zero ΔQ values, supporting the hypothesis that both years were influenced by relatively weak TME tracks. This result initially indicates that the model may be improved for over-prediction by adjusting the annual TME index to only consider TME tracks that meet a certain ΔQ threshold. However, early attempts at applying such a threshold led to under-prediction in other years, so a robust correction is not immediately evident. Further research is needed on this topic.

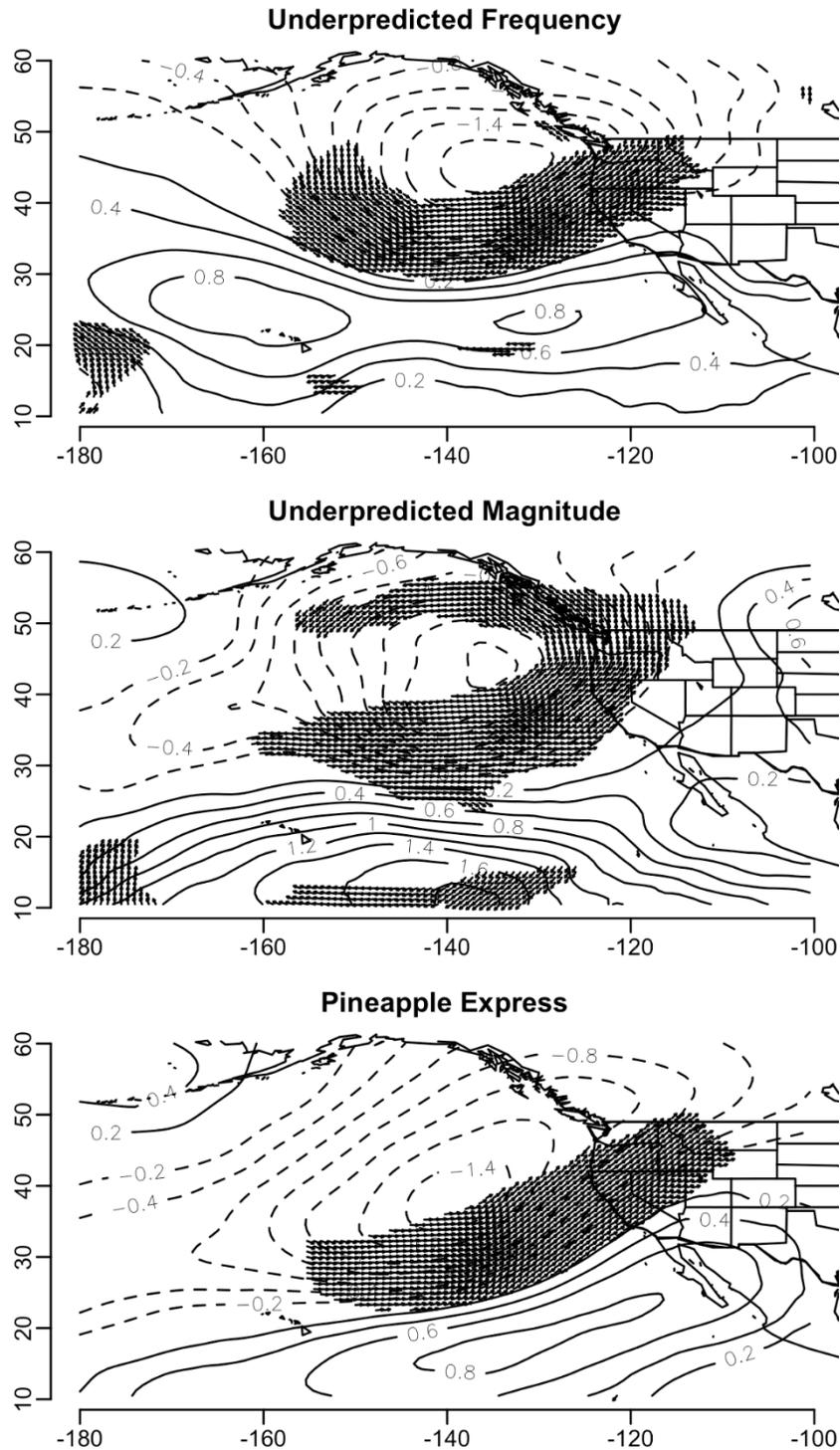


Figure 11. Anomaly composites of 500 mb level vector winds and geopotential heights during days with extreme rainfall in years with the worst Poisson model performance (top panel), the worst exponential model performance (middle panel), and the largest TME index (lower panel).

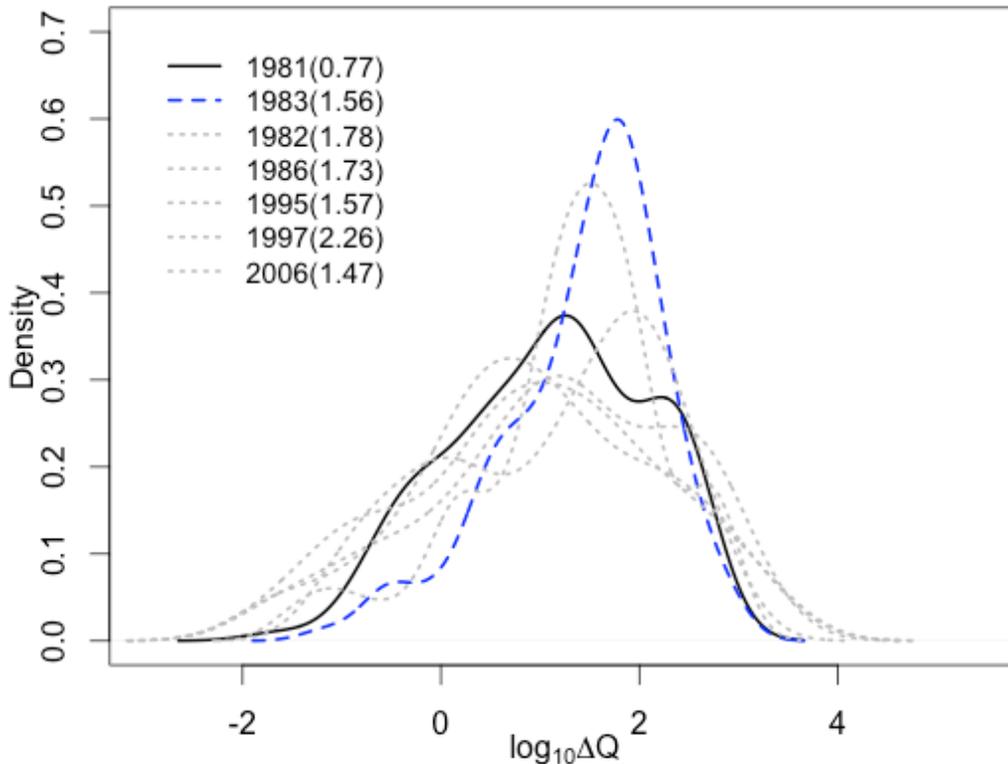


Figure 12. The distribution of the logarithm of ΔQ during years in which the frequency of rainfall extremes are over-predicted (1981 and 1983), as well other TME years. The TME annual index value for each year is given in parentheses in the legend.

5. Conclusion

Nearly 30 years ago, Hirschboeck [1988] proposed a systematic characterization of extreme floods based on the causal, climatological mechanisms that determine how storms vary across the United States. Despite this early work, engineering hydrology and insurance practices have tended to favor a statistical characterization of extreme events in lieu of a physical interpretation of their root causes. While this approach can be both simple and effective for planning purposes in a stationary world, the modern challenges of climate change and rapid urbanization, along with revelations regarding decadal and longer cycles in the climate system, lead us to question the adequacy of a strictly statistical approach to extreme event frequency analysis.

This paper presented a regional hierarchical Bayesian model for non-stationary precipitation extremes in Northern California that conditioned a POT model on the physical atmospheric mechanisms known to govern regional flooding, namely TMEs from the equatorial Pacific. The full Bayesian model outperformed other, simpler models that did not leverage the information content of the TME index, particular with respect to the frequency of extremes. Furthermore, the model was able to integrate TME-related information with site-specific characteristics to improve the regionalization of the POT model to ungauged locations. Still, the potential for substantial model improvement is large

if the time-invariant component of the Poisson model could be better regionalized across space.

A diagnostic analysis of years in which the hierarchical model most underperformed its simpler counterpart suggested that the time-varying component of the model could be most improved by including additional covariates that characterize other relevant flood-generating climate mechanisms, such as frontal systems. Even without this information, however, the inter-annual variability of the TME process was shown to cause substantial year-to-year modulations in extreme event risk over the region. Alternative realizations of the same TME process could lead to periods with long streaks of elevated extreme event risk that have implications for regional flood management strategies.

Looking forward, the integration of statistical models of extremes with a physical understanding of their causal mechanisms provides a viable way to link classical risk estimation approaches for water resources planning and management with state-of-the-art climate science and modeling. The regional model proposed in this work could be coupled with long-term projections of TME frequency and magnitude under anthropogenic forcing to provide a notable advance towards this end. Some recent work has focused on projections of the Pineapple Express [Dettinger et al., 2011] and Gulf Stream [Lavers et al., 2013] TME phenomena, but more research is needed to understand how credibly AOGCMs reproduce these atmospheric features and how this information can be downscaled and coupled with local observations for use in long-term planning. Correspondingly, the development of an index related to frontal storms or extratropical cyclones may add strength to the model developed here for this region.

We are pursuing extensions of such models to a multilevel structure where atmospheric circulation indices derived from physical considerations as well as indices for tropical ocean conditions can be used to predict the birth and evolution of TMEs, which in turn are used to predict extreme regional precipitation and flooding. The motivation for building such models is to provide a mechanism for the causal analysis of the processes and their comparison across climate re-analysis models and climate models used for seasonal forecasting and climate change projections. Such comparisons would allow a mechanistic assessment of what aspects of the causal chain to extreme precipitation are well modeled or deficient in the numerical integrations of the models.

Acknowledgements

This research was made possible with funding from the NOAA Climate and Global Change Fellowship Program. The second author was supported by an IPA from the US Army Corps of Engineers, by a grant from AI, and NOAA Grant NA10OAR4310159.

References

Alexander, L.V., P. Uotila, and N. Nicholls (2009), Influence of sea surface temperature variability on global temperature and precipitation extremes, *Journal of Geophysical Research*, 114, D18116, doi:10.1029/2009JD012301.

Berg, P., C. Moseley, and J. O. Haerter (2013), Strong increase in convective precipitation in response to higher temperatures, *Nature Geoscience*, 6(3), 181-185.

- Box, G. E. P., and G. Jenkins (1970), *Time Series Analysis, Forecasting and Control*, Holden-Day, Boca Raton, Fla.
- Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer, London.
- Davison, A.C., and R.L. Smith (1990), Models for Exceedances over High Thresholds, *Journal of the Royal Statistical Society Series B*, 52 (3), 393-442.
- Dee, D. P. et al. (2011), The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q.J.R. Meteorol. Soc.*, 137: 553–597. doi: 10.1002/qj.828
- Dettinger, M.D. (2013), Atmospheric rivers as drought busters on the U.S. West Coast, *Journal of Hydrometeorology*, 14, 1721-1732.
- Dettinger, M.D., and D.R. Cayan (2014), Drought and the California delta – a matter of extremes, *San Francisco Estuary and Watershed Science*, 12(2).
- Dettinger, M.D., F.M. Ralph, T. Das, P.J. Neiman, and D. Cayan (2011), Atmospheric rivers, floods, and the water resources of California, *Water*, 3, 455-478.
- Dirmeyer, P.A., and J.L. Kinter III (2009), The “Maya Express”: Floods in the U.S. Midwest, *Eos Transactions of the American Geophysical Union*, 90 (12), 101-108.
- Dirmeyer, P.A., and J.L. Kinter III (2010), Floods over the U.S. Midwest: A regional water cycle perspective, *Journal of Hydrometeorology*, 11, 1172-1181.
- Duane, S., A.D. Kennedy, B.J. Pendleton, and D. Roweth (1987), Hybrid Monte Carlo, *Physics Letters B*, 195 (2), 216-222.
- Ensor, L.A., and S.M. Robeson (2008), Statistical Characteristics of Daily Precipitation: Comparisons of Gridded and Point Datasets, *J. Appl. Meteor. Climatol.*, 47, 2468–2476. doi: <http://dx.doi.org/10.1175/2008JAMC1757.1>
- Fischer, E.M., U. Beyerle, and R. Knutti (2013), Robust spatially aggregated projections of climate extremes, *Nature Climate Change*, 3, 1033-1038.
- Friederichs, P., and T.L. Thorarinsdottir (2012), Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction, *Environmetrics*, 23 (7), 579-594.
- Gelman, A., and D. B. Rubin (1992), Inference from iterative simulation using multiple sequences, *Stat. Sci.*, 7(4), 457–511.
- Hersbach H. (2000), Decomposition of the continuous ranked probability score for ensemble prediction systems, *Weather Forecast*, 15, 559–570.
- Hirabayashi, Y. et al. (2013), Global flood risk under climate change, *Nature Climate Change Letters*, 3, doi: 10.1038/NCLIMATE1911.
- Hirschboeck, K. K.: Flood Hydroclimatology, in: *Flood Geomorphology*, edited by: Baker, V. R., Kochel, R. C., and Patton, P. C., JohnWiley & Sons, 27–49, 1988.
- Houze Jr., R.A., K. L. Rasmussen, S. Medina, S. R. Brodzik, and U. Romatschke, 2011: Anomalous Atmospheric Events Leading to the Summer 2010 Floods in Pakistan. *Bull. Amer. Meteor. Soc.*, 92, 291–298. doi: <http://dx.doi.org/10.1175/2010BAMS3173.1>
- Jain, S., and U. Lall (2000), Magnitude and timing of annual maximum floods: Trends and large-scale climate associations for the Blacksmith Fork River, Utah, *Water Resources Research*, 36, 3641-3651.

- Jain, S., and U. Lall (2001), Floods in a changing climate: Does the past represent the future? *Water Resources Research*, 37 (12), 3193-3205.
- Karamperidou, C., F. Cioffi, and U. Lall (2012), Surface temperature gradients as diagnostic indicators of midlatitude circulation dynamics, *J. Climate*, 25, 4154-4171.
- Khaliq, M.N., T.B.M.J. Ouarda, J.-C. Ondo, P. Gachon, and B. Bobee (2006), Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review, *Journal of Hydrology*, 329 (3-4), 534-552.
- Knippertz, P., and H. Wernli (2010), A Lagrangian climatology of tropical moisture exports to the northern hemispheric extratropics, *J. Climate*, 23, 987-1003.
- Knippertz, P., H. Wernli, and G. Glaser (2012), A global climatology of tropical moisture exports, *J. Climate*, 26, 3031-3045.
- Komori, D., S. Nakamura, M. Kiguchi, A. Nishijima, D. Yamazaki, S. Suzuki, A. Kawasaki, K. Oki, and T. Oki (2012), Characteristics of the 2011 Chao Phraya River flood in Central Thailand, *Hydrological Research Letters*, 6, 41-46.
- Koutsoyiannis, D. (2006), Nonstationarity versus scaling in hydrology, *Journal of Hydrology*, 324 (1-4), 239-254.
- Koutsoyiannis, D., and A. Montanari (2014), Negligent killing of scientific concepts: the stationarity case, *Hydrological Sciences Journal*, doi: 10.1080/02626667.2014.959959.
- Kwon, H.-H., C. Brown, and U. Lall (2008), Climate informed flood frequency analysis and prediction in Montana using hierarchical Bayesian modeling, *Geophysical Research Letters*, 35, L05404, doi:10.1029/2007GL032220.
- Lackmann, G.W. (2013), The south-central U.S. flood of May 2010: present and future, *Journal of Climate*, 26, 4688-4709.
- Laio, F., and S. Tamea (2007), Verification tools for probabilistic forecasts of continuous hydrologic variables, *Hydrol. Earth Syst. Sci.*, 11, 1267-1277.
- Lavers, D.A., and G. Villarini (2013), Atmospheric rivers and flooding over the central United States, *Journal of Climate*, 26, 7829-7836.
- Lavers, D.A., R.P. Allan, E.F. Wood, G. Villarini, D.J. Brayshaw, and A. J. Wade (2011), Winter floods in Britain are connected to atmospheric rivers, *Geophysical Research Letters*, 38, L23803, doi:10.1029/2011GL049783.
- Lavers, D.A., R.P. Allan, G. Villarini, B. Lloyd-Hughes, D.J. Brayshaw, and A.J. Wade (2013), *Environmental Research Letters*, 8, doi: 10.1088/1748-9326/8/3/034010.
- Lenderink, G., and E. van Meijgaard (2008), Increases in hourly precipitation extremes beyond expectations from temperature changes, *Nature Geoscience*, 1, 511-514.
- Lima, C.H.R., and U. Lall (2010), Spatial scaling in a changing climate: A hierarchical Bayesian model for non-stationary multi-site annual maximum and monthly streamflow, *J. Hydrol.*, 383, 307-318.

- Lin, C.A., L. Wen, G. Lu, Z. Wu, J. Zhang, Y. Yang, Y. Zhu, and L. Tong (2010), Real-time forecast of the 2005 and 2007 summer severe floods in the Huaihe River Basin of China, *Journal of Hydrology*, 381 (1-2), 33-41.
- Lu, M., U. Lall, A. Schwartz, and H.-H. Kwon (2013), Precipitation predictability associated with tropical moisture exports and circulation patterns for a major flood in France in 1995, *Water Resources Research*, 49, 6381-6392.
- Marsh, T.J. (2004), The January 2003 flood on the Thames, *Weather*, 59, 59-62.
- Milly, P.C.D., R.T. Wetherald, K.A. Dunne, and T.L. Delworth (2002), Increasing risk of great floods in a changing climate, *Nature*, 415, 514-517.
- Moore, B.J., P.J. Neiman, F.M. Ralph, and F.E. Barthold (2012), Physical processes associated with heavy flooding rainfall in Nashville, Tennessee, and vicinity during 1-2 May 2010: the role of an atmospheric river and mesoscale convective systems, *Monthly Weather Review*, 140, 358-378.
- Muller, C.J., P.A. O’Gorman, and L.E. Back (2011), Intensification of precipitation extremes with warming in a cloud resolving model, *J. Climate*, 24, 2784-2800.
- Munich Re (2012), *Natural Catastrophes 2011: Analyses, Assessments, Positions, Topics Geo*, Munich Re Reinsurance Company: Munich, Germany.
- Nakamura, J., U. Lall, Y. Kushnir, A. W. Robertson, and R. Seager (2013), Dynamical Structure of Extreme Floods in the U.S. Midwest and the United Kingdom. *J. Hydrometeorol*, 14, 485–504. doi: <http://dx.doi.org/10.1175/JHM-D-12-059.1>
- Neiman, P.J., F.M. Ralph, G.A. Wick, J.D. Lundquist, and M.D. Dettinger (2008), Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the west coast of North America based on eight years of SSM/I satellite observations, *Journal of Hydrometeorology*, 9, 22-47.
- Neiman, P.J., L.J. Schick, F. M. Ralph, M. Hughes, and G.A. Wick (2011), Flooding in western Washington, the connection to atmospheric rivers, *Journal of Hydrometeorology*, 12, 1337-1358.
- Olsen, J. R., Stedinger, J. R., Matalas, N. C. and Stakhiv, E. Z. (1999), Climate variability and flood frequency estimation for the Upper Mississippi and Lower Missouri rivers, *Journal of the American Water Resources Association*, 35, 1509–1523. doi: 10.1111/j.1752-1688.1999.tb04234.x
- Pickands, J. (1975), Statistical inference using extreme order statistics, *Annals of Statistics*, 3, 119–131.
- Ralph, F.M., T. Coleman, P.J. Neiman, R.J. Zamora, and M.D. Dettinger (2013), Observed impacts of duration and seasonality of atmospheric-river landfalls on soil moisture and runoff in coastal Northern California, *Journal of Hydrometeorology*, 14, 443-459.
- Romps, D.M. (2011), Response of tropical precipitation to global warming, *J. Atmos. Sci.*, 68, 123-138.

- Schneider, J.M., Ford Jr, D.L. (2011), Evaluating PRISM precipitation grid data as possible surrogates for station data at four sites in Oklahoma, Oklahoma Academy of Science Proceedings, 90, 77-88.
- Shane R.M., and W.R. Lynn (1964), Mathematical model for flood risk evaluation, J Hydraul Eng Div, 90(HY 6), 4119-4122.
- Slingo, J., Belcher, S., Scaife, A., McCarthy, M., Saulter, A., McBeath, K., Jenkins, A., Huntingford, C., Marsh, T., Hannaford, J., Parry, S. (2014), The Recent Storms and Floods in the UK, UK Met. Office and Centre for Ecology and Hydrology. <http://www.metoffice.gov.uk/research/news/2014/uk-storms-and-floods>.
- St. George, S., and T. R. Ault (2011), Is energetic decadal variability a stable feature of the central Pacific Coast's winter climate?, J. Geophys. Res., 116, D12102, doi:10.1029/2010JD015325.
- Ulbrich, U., T. Brucher, A.H. Fink, G.C. Leckebusch, A. Kruger, and J.G. Pinto (2003), The central European floods of August 2002: Part 1 - Rainfall periods and floods development, Weather, 58, 371-377.
- van den Honert, R.C., and J. McAneney (2011), The 2011 Brisbane Floods: Causes, Impacts and Implications. Water, 3, 1149-1173.
- Villarini, G., F. Serinaldi, J. A. Smith, and W. F. Krajewski (2009), On the stationarity of annual flood peaks in the continental United States during the 20th century, Water Resour. Res., 45, W08417, doi:10.1029/2008WR007645.
- Ward, P.J., W. Beets, L.M. Bouwer, J.C.J.H. Aerts, and H. Renssen (2010), Sensitivity of river discharge to ENSO, Geophysical Research Letters, 37, L12402, doi:10.1029/2010GL043215.
- Weigel, A., M. Liniger, and C. Appenzeller (2007), The discrete Brier and ranked probability skill scores, Monthly Weather Review, 135(1), 118-124.
- Wernli, H., and H. C. Davies, (1997), A Lagrangian-based analysis of extratropical cyclones. I: The method and some applications, Quart. J. Roy. Meteor. Soc., 123, 467-489.
- Zhu, Y., and R.E. Newell (1994), Atmospheric rivers and bombs, Geophysical Research Letters, 21, 1999-2002.
- Zhu, Y., and R.E. Newell (1998), A proposed algorithm for moisture fluxes from atmospheric rivers, Monthly Weather Review, 126, 725-735.
- Zong, Y., and X. Chen (2000), The 1998 Flood on the Yangtze, China, Nat. Hazards, 22, 165-184.